Chapter 2—Summarizing and Graphing Data

Important Characteristics of Data

1. Center—the average or middle value of a data set
2. Variation—a measure of the variation of the data values
3. Distribution—the shape of the distribution
4. Outliers—extremely high or low values in a data set
5. Time—changing characteristics of the data over time

Descriptive statistics—used to summarize or describe a data set

Inferential statistics—used when we take a sample and use that sample to draw conclusions about some population

Section 2-2: Frequency Distributions

Frequency distribution—lists data values and their corresponding frequencies

Example: The following is a frequency distribution of the vacation days for people in a particular company.

Days of vacation	Frequency
0-3	22
4-7	17
8-11	14
12-15	20
16-19	27

There are five classes.

Lower class limits—the smallest numbers that can belong to the difference classes (0 is the lower class limit for the first class, 4 is the lower class limit for the second class, etc.)

Upper class limits—the largest numbers that can belong to the different classes (3 is the upper class limit for the first class, 7 is the upper class limit for the second class, etc.)

Class boundaries—the numbers used to separate classes, but without the gaps created by class limits. They fill the gaps between classes by splitting the difference between the end of one class and beginning of another class.

Class midpoints—the average of the lower and upper class limits for each class

Class width—the difference between two consecutive lower class limits or two consecutive lower class boundaries (the class width is 4).

Example: Identify the class width, the class midpoints, and class boundaries for the following frequency distribution.

Blood Pressure	Frequency
80-99	9
100-119	24
120 – 139	5
140 – 159	1
160 – 179	0
180 – 199	1
Total	40

Class width = 100-80 = 20
Class midpoints = 89.5, 109.5, 129.5, 149.5, 169.5, 189.5
Class boundaries = 79.5, 99.5, 119.5, 139.5, 159.5, 179.5, 199.5

Cumulative frequency—the sum of the frequencies for that class and all previous classes

Relative frequency = (class frequency)/(sum of all frequencies)

Example: For the previous problem, find the cumulative frequency and the relative frequency.

Blood Pressure	Frequency	Relative Frequency	Cumulative Frequency
80 – 99	9	.225	9
100 – 119	24	.600	33
120 – 139	5	.125	38
140 – 159	1	.025	39
160 – 179	0.	.000	39
180 – 199	1	.025	40

Example: On a math test, the scores of 24 students were
97 78 77 68 77 77 97 87 60 61 87 78 78 87 77 78 87 77 78 87 78 87 87 69
Construct a frequency distribution using four classes.

To construct a frequency distribution

1. Decide on how many classes you want--4.
2. Calculate the class width.
   Class width = (highest value – lowest value)/number of classes, and then round up.
   Class width = (97-60)/4 = 9.25, round up to 10
3. Choose the lowest value, or a convenient value that is a little smaller, as the lower class limit of the first class. 60
4. Using the lower limit of the first class and the class width, list the other lower class limits. 60, 70, 80, 90
5. List the upper class limits. 69, 79, 89, 99
6. Put a tally mark the appropriate class for each data value. Use the tally marks to find the frequency for each class.

Score	Tally	Frequency	Cumulative Frequency	Relative Frequency
60-69		4	4	.17
70-79		11	15	.46
80-89		7	22	.29
90-99		2	2	.08

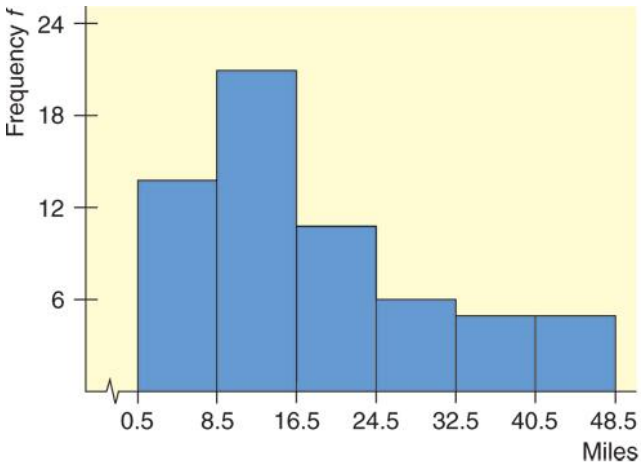
Section 2-3: Histograms

Displaying **quantitative** data (numbers)

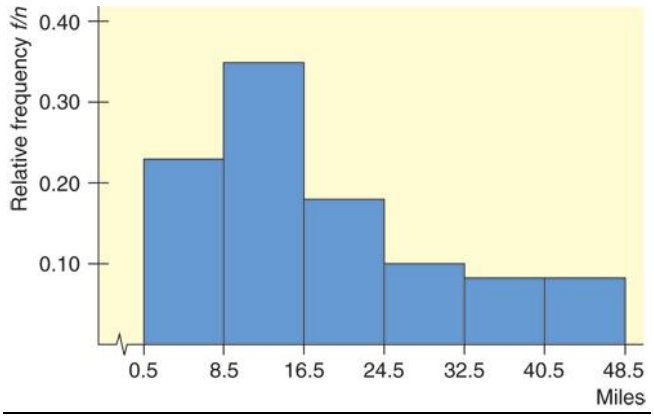
Histogram—a bar graph in which the horizontal scale represents classes of data values and the vertical scale represents frequencies. The heights of the bars correspond to frequency values, and the bars are drawn adjacent to each other.

A histogram is used to display quantitative data (numbers).

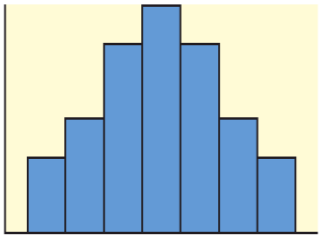
Example: The following is a histogram that displays commuting distances for a sample of drivers.



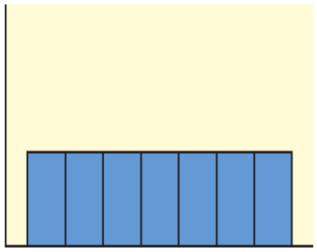
Relative frequency histogram—the vertical scale is marked with relative frequencies instead of frequencies



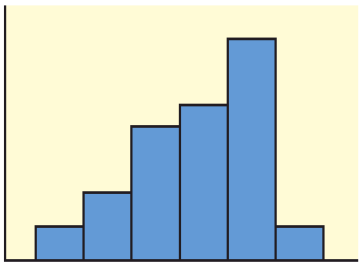
Common Shapes of Histograms



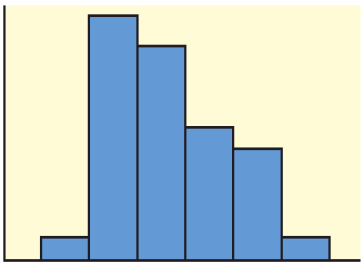
Symmetrical



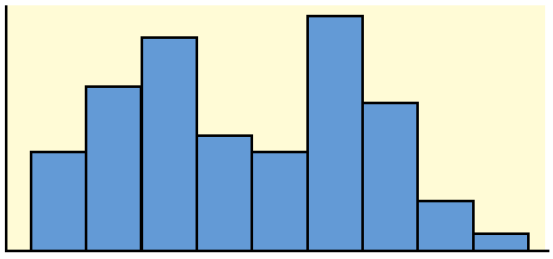
Uniform



Skewed left



Skewed right



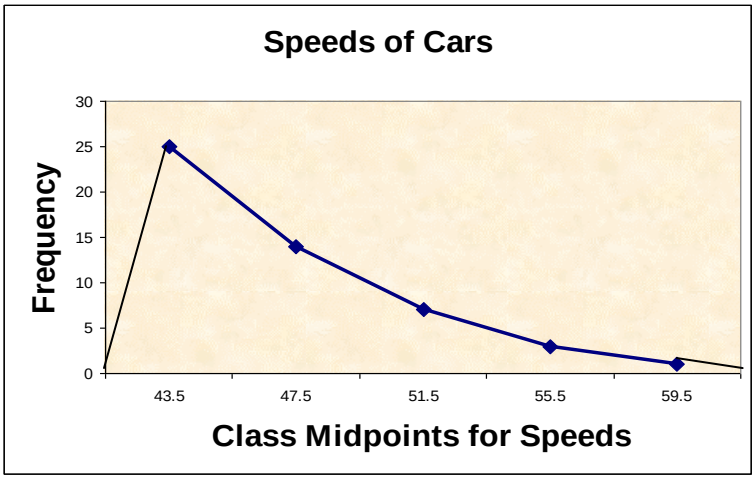
Bimodal

Section 2-4: Statistical Graphics

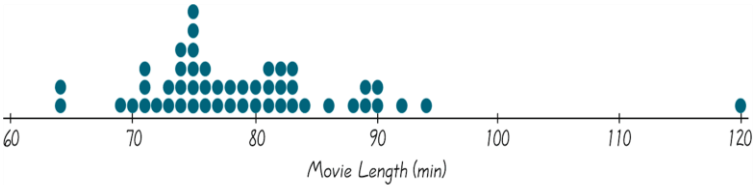
Frequency polygon—uses line segments connected to points located directly above class midpoint values.

Example: The following table contains the speeds of cars passing along a stretch of road.

Speed	Frequency	Class Midpoint
42-45	25	43.5
46-49	14	47.5
50-53	7	51.5
54-57	3	55.5
58-61	1	59.5



Dotplots – a graph in which each data value is plotted as a point (or dot) along a scale of values.



Stem-and-leaf plots—a graph that separates each data value into a "stem" and a "leaf"

Example: Create a stem and leaf plot for the following ages:
31, 32, 32, 32, 33, 35, 36, 36, 37, 37, 38, 39, 39, 40, 40, 40, 41, 42, 42, 43, 43, 44, 45, 45, 46, 46, 47, 48, 48, 51, 53, 55, 56, 56, 60, 60, 61, 62, 76

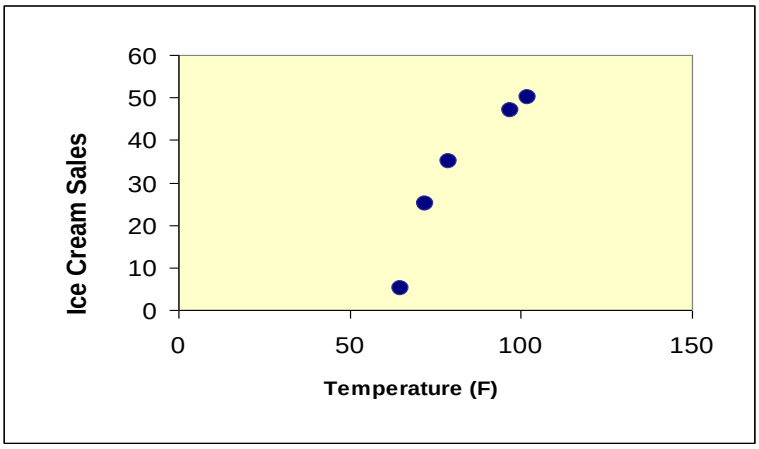
Stem (tens)	Leaves (units)
3	1 2 2 2 3 5 6 6 7 7 8 9 9
4	0 0 0 1 2 2 3 3 4 5 5 6 6 7 8 8
5	1 3 5 6 6
6	0 0 1 2
7	6

Scatter plots—a plot of paired (x,y) data with a horizontal x-axis and a vertical y-axis

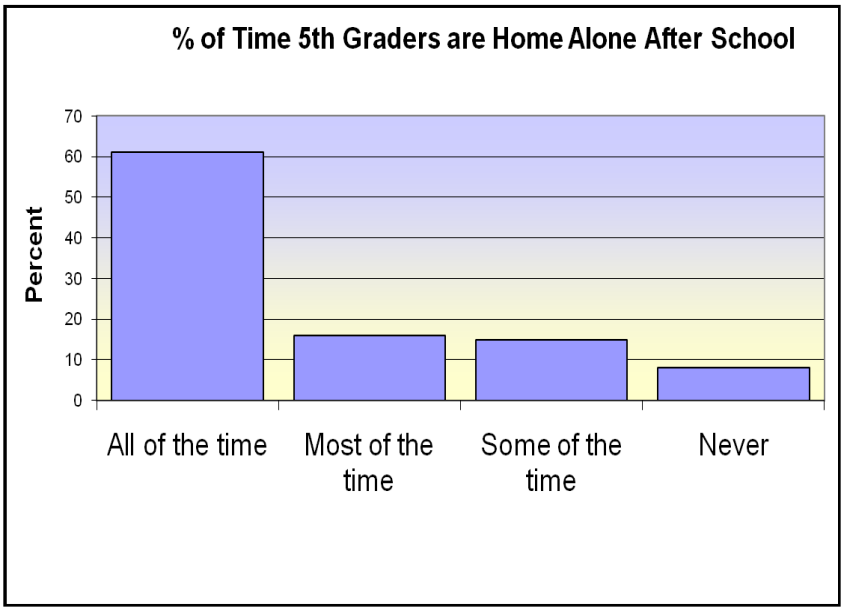
The pattern of the data is helpful in determining whether there is some relationship between two variables.

Temperature	72	79	65	97
Ice Cream Sales	25	35	5	47

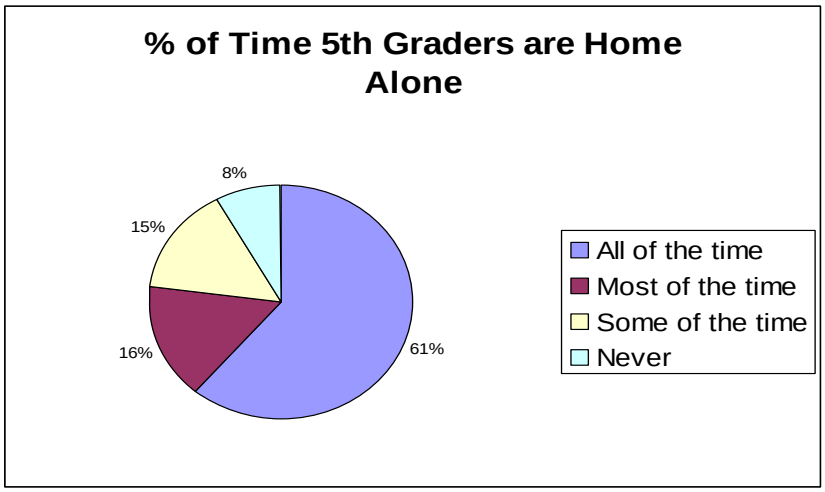
Displaying **qualitative** data (non-numbers)



Pareto chart—a bar graph for qualitative data in which the vertical scale can represent frequencies or relative frequencies. The tallest bar is on the left, and the smaller bars are farther to the right.



Pie charts—depicts qualitative data as slices of a pie.



Time series graph—data that have been collected at different points in time